

University of Groningen

Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service

Mohammadi, A.; Salehi-Rad, M. R.; Wit, E. C.

Published in:
Computational Statistics

DOI:
[10.1007/s00180-012-0323-3](https://doi.org/10.1007/s00180-012-0323-3)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Mohammadi, A., Salehi-Rad, M. R., & Wit, E. C. (2013). Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service. *Computational Statistics*, 28(2), 683-700.
<https://doi.org/10.1007/s00180-012-0323-3>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service

A. Mohammadi · M. R. Salehi-Rad · E. C. Wit

Received: 3 July 2010 / Accepted: 21 March 2012 / Published online: 7 April 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The paper proposes Bayesian framework in an M/G/1 queuing system with optional second service. The semi-parametric model based on a finite mixture of Gamma distributions is considered to approximate both the general service and re-service times densities in this queuing system. A Bayesian procedure based on birth-death MCMC methodology is proposed to estimate system parameters, predictive densities and some performance measures related to this queuing system such as stationary system size and waiting time. The approach is illustrated with several numerical examples based on various simulation studies.

Keywords Gamma mixtures · Bayesian inference · MCMC · Birth-death predictive distribution · M/G/1 queue · Optional service

1 Introduction

Bayesian analysis of queuing systems is a relatively recent research area. Some useful references are [Armero and Bayarri \(1994a,b\)](#), [Rios et al. \(1998\)](#), [Ausin et al. \(2004, 2007\)](#), [Ausin and Wiper \(2007\)](#), and [Ramirez et al. \(2008\)](#). The queuing systems that they considered customers depart the system after taking their service. But in many

A. Mohammadi (✉) · E. C. Wit
Johann Bernoulli Institute, University of Groningen, Groningen, The Netherlands
e-mail: a.mohammadi@rug.nl

E. C. Wit
e-mail: e.c.wit@rug.nl

M. R. Salehi-Rad
Department of statistics, Allameh Tabataba'i University, Tehran, Iran
e-mail: salehirad@atu.ac.ir

applied queuing systems, some customers need to be re-serviced after taking their main service. For example, in a production line, some items might fail and require repair. In these kinds of problems, we must re-service some items.

The primary aim of this paper is to propose a Bayesian inference scheme for an M/G/1 queuing system in which some customers with probability p need re-servicing. This queuing system has a service unit, in which customers arrive according to a Poisson process and demanding service with a general distribution. A fraction p of these customers request re-service with possibly another general distribution. From a classical queuing theory perspective, this queuing system has been studied by Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004); they considered three alternatives for re-servicing in this queuing system and obtained the mean busy period, the probability of the idle period and the probability generating function (pgf) of the steady-state system size. More recently, this queuing system has been studied by Mohammadi and Salehi-Rad (2012) based on Bayesian approach by using a mixture of truncated Normal distributions.

The main contribution of this paper is to introduce a semi-parametric model for the general density of service and re-service based on a mixture of Gamma distributions, providing an alternative Bayesian approach for approximating the general distributions in queuing systems based on former work. Secondly, we will introduce a Bayesian algorithm based on the birth-death MCMC approach of Stephens (2000a) in order to fit this model to data.

The use of finite mixture distributions is very common and the Bayesian approach provides an important tool in semi-parametric density estimation, see for instance Diebolt and Robert (1994) and Robert (1996). Recently, MCMC methods for fully Bayesian mixture models of unknown dimension have been proposed; see Frühwirth-Schnatter (2006). Green (1995) introduced the reversible jump technique (RJ-MCMC) and Richardson and Green (1997) used this methodology to analyze Normal mixtures. This type of algorithm was used by Rios et al. (1998) for mixture of Exponential distributions, Wiper et al. (2001) for mixture of Gamma distributions and Ausin et al. (2004) for mixture of Erlang distributions. More recently, in the context with this methodology, Stephens (2000a) rekindled interest in the use of continuous time birth-death methodology (BD-MCMC) for variable dimension problems. This type of methodology was used by Ausin and Wiper (2007) for mixture of Erlang distributions and Ramirez et al. (2008) for the mixture of Pareto distributions and Mohammadi and Salehi-Rad (2012) for the mixture of truncated Normal distributions. Moreover, Cappé et al. (2003) investigated the similarity between the reversible jump and birth-death methodology.

The paper is structured as follows. Section 2 illustrates the M/G/1 queuing system with optional second service where we consider a mixture of Gamma distributions to approximate the general densities of service and re-service times. Then, we use some results obtained by Salehi-Rad and Mengersen (2002), Salehi-Rad et al. (2004) and Mohammadi and Salehi-Rad (2012) which allow us to estimate the mean number of customers in the system, mean busy period and probability of the idle period for this queuing system. Section 3 describes our Bayesian approach by defining prior distributions, obtaining posterior conditional distributions and propose a birth-death MCMC algorithm to obtain a sample from the posterior distributions of the parameters

of the predictive service and re-service times distributions. Section 4 explains how to approximate the general densities of service and re-service times by using the data generated from the birth-death MCMC algorithm. Section 5 demonstrates how to estimate the system parameters and some performance measures of our queuing system from the BD-MCMC output. Section 6 illustrates our methodology by performing several simulation studies. The paper concludes with a discussion of the relevance of the various extensions in Sect. 7.

2 The M/G/1 queuing system with optional second service

Throughout, we are considering an M/G/1 queue in which some customers with probability p need re-service, with First Come First Serve discipline, and independence between inter-arrival and service times. In this queuing system failed items are stockpiled in a failed queue (FQ) and re-served only after all customers in main queue (MQ) are serviced. After completion of re-service of all items in FQ, the server returns to MQ if there are any customers waiting in MQ; otherwise, the server is idle.

The variable T is the inter-arrival time with an exponential distribution. For service times, we suppose that service (S) and re-service (\tilde{S}) times are independent and have general distributions, denoted by $B_1(\cdot)$ and $B_2(\cdot)$ with means μ_1, μ_2 and variances δ_1, δ_2 respectively. For these general distributions, we need a model flexible enough to deal with typical features in service and re-service time distributions (skewness, multimodality, lots of mass near zero, even possibly a mode) and permits usual computations in queuing applications. Thus, we propose a semi-parametric model based on a mixture of Gamma distributions. If S is a service time, we assume

$$B_1\left(s \left| \begin{matrix} \theta \\ \sim_1 \end{matrix} \right. \right) = \sum_{i=1}^{k_1} \pi_{1i} G(s | \alpha_{1i}, \beta_{1i}), \quad s > 0$$

where $\theta_{\sim_1} = \left(k_1, \pi_{\sim_1}, \alpha_{\sim_1}, \beta_{\sim_1} \right)$, k_1 is the unknown number of mixture components, $\pi_{\sim_1} = (\pi_{11}, \pi_{12}, \dots, \pi_{1k_1})$ are weights and $G(s | \alpha_{1i}, \beta_{1i})$ represents the Gamma density function, for $i = 1, \dots, k_1$, that is,

$$G(s | \alpha_{1i}, \beta_{1i}) = \frac{(\beta_{1i})^{\alpha_{1i}}}{\Gamma(\alpha_{1i})} s^{\alpha_{1i}-1} e^{-\beta_{1i}s}, \quad s > 0.$$

Likewise, if \tilde{S} is a re-service time, we have

$$B_2\left(\tilde{s} \left| \begin{matrix} \theta \\ \sim_2 \end{matrix} \right. \right) = \sum_{i=1}^{k_2} \pi_{2i} G(\tilde{s} | \alpha_{2i}, \beta_{2i}), \quad \tilde{s} > 0$$

where $\theta_{\sim_2} = \left(k_2, \pi_{\sim_2}, \alpha_{\sim_2}, \beta_{\sim_2} \right)$, k_2 and $\pi_{\sim_2} = (\pi_{21}, \pi_{22}, \dots, \pi_{2k_2})$ have the same interpretation as for the service times density.

Thus, we have a queuing system with two queues, main queue and failed queue, and one server. Therefore, all parameters of these queuing system are $\left(\lambda, \theta_1, \theta_2, p\right)$, in which λ is the parameter of inter-arrival times and p is the probability of the items needing re-service.

2.1 Some performance measures in this queuing system

We assume that the queuing system is in equilibrium. This assumption is equivalent with assuming that the traffic intensity, ρ , is less than one (Medhi 1982). For our queuing system $\rho = \rho_1 + p\rho_2$, in which $\rho_1 = \lambda\mu_1$ is traffic intensity in MQ and $\rho_2 = \lambda\mu_2$ is traffic intensity in FQ. As a result

$$\rho = \lambda(\mu_1 + p\mu_2). \quad (1)$$

Under this steady state condition, other performance measures will be obtained.

2.1.1 The expectation of busy period and probability of idle period

To acquire the expectation of busy period for this queuing system we have

$$E[\text{busy period}] = E[\text{busy period in MQ}] + pE[\text{busy period in FQ}],$$

which the first expressions is equal to $\mu_1/(1 - \lambda\mu_1)$ and the second is $p\mu_2/(1 - \lambda\mu_1)$. Therefore, by some computation

$$E[\text{busy period}] = \frac{\mu_1 + p^2\mu_2}{1 - \lambda\mu_1}. \quad (2)$$

Furthermore, the probability of idle period is

$$\text{Pr}(\text{idle period}) = \frac{1 - \lambda\mu_1}{1 + p^2\lambda\mu_2}. \quad (3)$$

For more details see Salehi-Rad and Mengersen (2002) and Salehi-Rad et al. (2004).

2.1.2 The expectation of the system size

For our queuing system, suppose that X_n is the number of customers remaining in MQ at the completion of the n th customer's service time also Y_n is the number of customers remaining in FQ at the completion of the n th customer's service time in the steady state. Mohammadi and Salehi-Rad (2012), by using the joint probability general function of (X_n, Y_n) obtained the following expression of the mean system size.

Theorem 2.1 (Mohammadi and Salehi-Rad 2012) *Mean number of customers in MQ and FQ are as below*

i)

$$E(X_n) = \rho_1 + \frac{\lambda^2 \delta_1 + \rho_1^2}{2(1 - \rho_1)} + \frac{p \left[\lambda^2 \delta_2 + \rho_2^2 + p \rho_2^2 \left(\frac{\lambda^2 \delta_1 + \rho_1(2 - \rho_1)}{(1 - \rho_1)} \right) \right]}{2(p \rho_2 + (1 - \rho_1) \Psi(1 - p + p B_2^*(\lambda)))} \quad (4)$$

ii)

$$E(Y_n) = \frac{p(2(1 - \rho_1)^2 + \lambda^2 \delta_1 + \rho_1(1 - \rho_1))}{2(p \rho_2 + (1 - \rho_1) \Psi(1 - p + p B_2^*(\lambda)))} \quad (5)$$

respectively, where

$$\Psi(u) = u B_1^*[\lambda(1 - \Psi(u))],$$

$B_1^*(.)$ and $B_2^*(.)$ are the Laplace Stieltjes Transform (LST) of the service and re-service times density, respectively.

According to mixture of Gamma distributions for service and re-service times, the LST of service and re-service times are given by

$$B_j^*(t) = \sum_{i=1}^{k_j} \pi_{ji} \left(\frac{\alpha_{ji}}{t + \beta_{ji}} \right)^{\alpha_{ji}}, \quad j = 1, 2$$

and the variance of service and re-service times are given by

$$\delta_j = \sum_{i=1}^{k_j} \pi_{ji}^2 \left(\frac{\alpha_{ji}}{\beta_{ji}^2} \right), \quad j = 1, 2.$$

3 Bayesian inference

In this section we propose the Bayesian approach to infer the system parameters $(\lambda, \theta_{\sim 1}, \theta_{\sim 2}, p)$. We observe n_t inter-arrival times $\tilde{t} = \{t_i\}_{i=1}^{n_t}$, n_{s1} service times $\tilde{s}_1 = \{s_{1i}\}_{i=1}^{n_{s1}}$, n_{s2} re-service times $\tilde{s}_2 = \{s_{2i}\}_{i=1}^{n_{s2}}$ and n_p indicators $\tilde{u} = \{u_i\}_{i=1}^{n_p}$, in which $u_i = 1$ if customer need re-service and $u_i = 0$ if customer does not need re-service. We assume independence between the arrival, service times, re-service times and the probability of re-service.

For the arrival rate, λ , we assume a Gamma prior distribution, $\lambda \sim G(\xi, \psi)$. Conditional on arrival data, the posterior distribution of λ is also Gamma distributed as $G(\xi + n_t, \psi + \sum_{i=1}^{n_t} t_i)$.

For the parameter p , we assume a Beta prior distribution, $p \sim \text{Beta}(a, b)$. The posterior distribution of p given \tilde{u} is Beta $(a + \sum_{i=1}^{n_p} u_i, b + n_p - \sum_{i=1}^{n_p} u_i)$.

In the following section, we propose a Bayesian framework for mixture of Gamma distributions to approximate the general distribution of the service and re-service times, $B_1(\cdot)$ and $B_2(\cdot)$.

3.1 Bayesian inference for mixture of Gamma distributions

To determine a Bayesian framework for the general distributions $B_1(\cdot)$ and $B_2(\cdot)$ based on mixture of Gamma distribution we assume that

$$B\left(s \mid \underset{\sim}{\theta}\right) = \sum_{i=1}^k \pi_i G\left(s \mid \alpha_i, \beta_i\right), \quad s > 0$$

where $\underset{\sim}{\theta} = \left(k, \underset{\sim}{\pi}, \underset{\sim}{\alpha}, \underset{\sim}{\beta}\right)$. First, as is usually done in mixture models (e.g. [Diebolt and Robert 1994](#)), we use a data augmentation algorithm, introducing for each datum, S_j , component indicator variables, Z_j , such that

$$P\left(Z_j = i \mid k, \underset{\sim}{\pi}\right) = \pi_i, \quad i = 1, \dots, k$$

Then, the conditional service time density, S_j , given that Z_j , is

$$S_j \mid Z_j = i \sim G\left(s_j \mid \alpha_i, \beta_i\right), \quad j = 1, \dots, n_s.$$

Following [Richardson and Green \(1997\)](#), we assume that the joint prior distribution on the mixture Gamma parameters, $\underset{\sim}{\theta} = \left(k, \underset{\sim}{\pi}, \underset{\sim}{\alpha}, \underset{\sim}{\beta}\right)$ can be factorized as

$$f(k, \underset{\sim}{\pi}, \underset{\sim}{\alpha}, \underset{\sim}{\beta}, \underset{\sim}{z}) \propto f(k) f(\underset{\sim}{\pi} \mid k) f(\underset{\sim}{z} \mid \underset{\sim}{\pi}, k) f(\underset{\sim}{\alpha} \mid k) f(\underset{\sim}{\beta} \mid k).$$

To determine the prior distributions for the parameters of mixture distribution, first, we assume a truncated Poisson distribution for the mixture size, k , as below

$$P(K = k) \propto \frac{\gamma^k}{k!}, \quad k = 1, \dots, k_{\max} \quad (6)$$

We define prior distributions for remaining parameters given that k , as below

$$\underset{\sim}{\pi} \mid k \sim D(\phi_1, \dots, \phi_k) \quad (7)$$

$$\alpha_i \mid k \sim G(v, v), \quad i = 1, \dots, k \quad (8)$$

$$\beta_i \mid k \sim G(\eta, \tau), \quad i = 1, \dots, k \quad (9)$$

where $D(\phi_1, \dots, \phi_k)$ denotes a Dirichlet distribution with parameters $\phi_i > 0$. Typically, we might set $\phi_i = 1$, for all $i = 1, \dots, k$, giving a uniform $U(0,1)$ prior for the weights.

Given k and the data, \tilde{s} , then it is straightforward to show the required posterior conditional distributions for the MCMC algorithm are as below

$$\begin{aligned} P\left(Z_j = i \mid \tilde{s}, k, \tilde{\pi}, \tilde{\alpha}, \tilde{\beta}\right) &\propto \pi_i G(s_j \mid \alpha_i, \beta_i), \quad i = 1, \dots, k, \\ \tilde{\pi} \mid \tilde{s}, \tilde{z}, k &\sim D(\phi_1 + n_1, \dots, \phi_k + n_k) \\ \tilde{\beta}_i \mid \tilde{s}, \tilde{z}, k &\sim G\left(\eta + n_i \alpha_i, \tau + \sum_{j: z_j = i} s_j\right), \quad i = 1, \dots, k \\ f\left(\alpha_i \mid \tilde{s}, \tilde{z}, k, \tilde{\beta}\right) &\propto \left(\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}\right)^{n_i} \left(\prod_{j: z_j = i} s_j\right)^{\alpha_i} \alpha_i^{\nu-1} e^{-\nu \alpha_i}, \end{aligned} \quad (10)$$

where $n_i = \#\{z_j = i\}$ for $i = 1, \dots, k$.

This mixture model is invariant to permutation of the labels $i = 1, \dots, k$. For identifiability, it is important to adopt a unique labeling. Unless stated otherwise, we use that the π_i are increasing; thus the prior distributions of the parameters are $k!$ times the product of the individual Gamma densities, restricted to the set $\pi_1 < \dots < \pi_k$, for more details see [Stephens \(2000b\)](#) and [Sperrin et al. \(2010\)](#).

In order to sample the posterior distributions, there are two main approaches in the context of mixture modeling with an unknown number of components: one approach is [Green's \(1995\)](#) reversible jump MCMC (RJ-MCMC) methodology. Another alternative approach is [Stephens \(2000a\)](#) birth-death MCMC (BD-MCMC) methodology. [Stephens \(2000a\)](#) introduced continuous time birth-death MCMC processes for variable dimension problems. [Cappé et al. \(2003\)](#) showed that the essential mechanism in this approach was the same as with RJ-MCMC algorithm. Here we apply the BD-MCMC, which is simpler to implement and we have found to give better results in practice. We briefly outline this algorithm in the following section. For more details, including details of the construction of the BD-MCMC methodology see [Stephens \(2000a\)](#) and [Frühwirth-Schnatter \(2006\)](#).

3.2 BD-MCMC algorithm

In this subsection, we obtain a sample from the posterior distributions of the parameters, $\tilde{\theta} = \left(k, \tilde{\pi}, \tilde{\alpha}, \tilde{\beta}\right)$, by a BD-MCMC algorithm. This algorithm is based on a birth-death process and was introduced by [Stephens \(2000a\)](#) in the context of Normal mixtures. With this approach, the model parameters are interpreted as observations from a marked point process and the mixture size, k , changes so that births and deaths of the mixture components occur in continuous time. The rates at which this happens determine the stationary distribution of the process.

In the birth-death process, births of mixture components occur at a constant rate which we might set equal to the parameter, γ , from the prior distribution of k in (6).

A birth increases the number of mixture components by one. Whenever a new component is born, its weight is generated from a Beta distribution with parameters $(1, k)$ and the remaining parameters are sampled from their posterior distributions. To include the new component, the old component weights are scaled down proportionally to make all the weights, including the new one, sum to 1, i.e. $\pi_i := \pi_i / (1 + \pi)$. The death rate of every mixture component is a likelihood ratio of the model with and without this component, given by

$$\Delta_j = \prod_{r=1}^{n_s} \left(\frac{B(s_r) - \pi_j g(s_r | \alpha_j, \beta_j)}{(1 - \pi_j) B(s_r)} \right), \quad j = 1, \dots, k \quad (11)$$

The total death rate, $\Delta = \sum_j \Delta_j$, of the process at any time is the sum of the individual death rates. A death decreases the number of mixture components by one. The birth and death processes are independent Poisson processes. Thus, the time of birth/death event is exponentially distributed with mean $1/(\Delta + \gamma)$. Therefore, a birth or death occurs with probabilities proportional to γ and Δ , respectively. With this explanation, we define an BD-MCMC algorithm based on [Stephens \(2000a\)](#) as follows.

Algorithm 3.1 Starting with initial values $k^{(0)}, \pi^{(0)}, \alpha^{(0)}$ and $\beta^{(0)}$, iterate the following steps:

1. Run the birth-death process for a fixed time t_0 and the birth rate γ
 - 1.1. Compute the death rates for each component, Δ_j , and the total death rate, $\Delta = \sum_j \Delta_j$
 - 1.2. Simulate the time to the next jump from an exponential distribution with mean $1/(\Delta + \gamma)$
 - 1.3. If the run time is less than t_0 continue otherwise proceed with step 2
 - 1.4. Simulate the type of jump: birth or death with probabilities

$$\Pr(\text{birth}) = \frac{\gamma}{\gamma + \Delta}, \quad \Pr(\text{death}) = \frac{\Delta}{\gamma + \Delta}$$
 - 1.5. Adjust the mixture components

MCMC steps conditional on k

2. Update the latent variables by sampling from $z^{(i+1)} \sim z \left| s, k^{(i+1)}, \pi^{(i)}, \alpha^{(i)}, \beta^{(i)} \right.$
 3. Update the weights by sampling from $\pi^{(i+1)} \sim \pi \left| s, k^{(i+1)}, z^{(i+1)} \right.$
 4. for $r = 1, \dots, k^{(i+1)}$
 - 4.1. Update β_r by sampling from $\beta_r^{(i+1)} \sim \beta_r \left| s, k^{(i+1)}, z^{(i+1)} \right.$
 - 4.2. Update α_r using a Metropolis-Hastings
 5. Set $i = i + 1$ and go to step 1.
-

Step one of the algorithm is the birth-death process described above. Following [Stephens \(2000a\)](#), we have chosen $t_0 = 1$ and a birth rate equal to the parameter, γ . As expected, we have found in practice that larger values of the birth rate produce better mixing but require more time in the computation of the algorithm.

Steps 2, 3 and 4.1 are standard Gibbs sampling moves, whereby the model parameters are updated conditional on the mixture size, k . The only complicated is step 4.2, where we introduce a Metropolis-Hastings step ([Hastings 1970](#)), to sample from the posterior distribution of α_i . From the shape of the target distribution, (10), we propose

a Gamma distribution with parameters $G(\nu, \nu)$. With this proposal distribution the acceptance probability for a candidate point, α^* , becomes

$$\Pr(\alpha_r, \alpha^*) = \min \left\{ 1, \left(\frac{\Gamma(\alpha_r)}{\Gamma(\alpha^*)} \right)^{n_r} \left(\prod_{j: z_j=r} \beta_r s_j \right)^{\alpha^* - \alpha_r} \right\}$$

Remark 3.1 Due to the overall similarity of the shape of the proposal distribution and the target distribution, the acceptance probability is much better – in simulation studies, Sect. 6, we get acceptance probabilities of around 20 % compared to 1 % elsewhere – than previous work (Wiper et al. 2001; Ausin et al. 2004).

Algorithm 3.1 produces a sample from the posterior distributions. Thus, we can run this BD-MCMC algorithm for the parameters of service and re-service time densities,

$$\theta_{\sim 1} = \left(k_1, \pi_{\sim 1}, \alpha_{\sim 1}, \beta_{\sim 1} \right), \theta_{\sim 2} = \left(k_2, \pi_{\sim 2}, \alpha_{\sim 2}, \beta_{\sim 2} \right), \text{ respectively.}$$

3.3 Model identification

The parameter k_1 and k_2 are model parameters, identifying models of a particular complexity. In this subsection, we discuss briefly how to perform “ideal Bayesian model identification” and a comparison with other approaches. There are many model selection criteria such as AIC, BIC, DIC, DIC+, MDL, Bayesian p values and posterior predictive checks (for more details see Claeskens and Hjort 2008). But most of them are either unsuitable for mixture models or complex (Celeux et al. 2006). This diversity of approaches, especially for variable-dimension parameters, reflects the different flavours of the model determination question that statisticians face.

In reality there are a number of reasons why this simple idealized view fails to reflect practical applications. We briefly describe some fundamental issues that face the practitioner wishing to perform model choice for a real Bayesian problem. We omit further details as they are covered in Green (2003) and Hastie and Green (2012).

First at all, prior model probabilities may be fictional: the ideal Bayesian has real prior probabilities reflecting scientific judgment or belief across the model space. In practice, however, such priors may not be commonly available. Secondly, Bayesian models have no chance of passing the test of a sensitivity analysis: in ordinary parametric problems we commonly find that inferences are rather insensitive to moderately large variations in prior assumptions, except when there are very few data. In fact, the opposite case, of high sensitivity, poses a greater challenge to the non-Bayesian as perhaps the data carry less information than hoped. Moreover, there may be improper parameter prior problems: in ordinary parametric problems it is commonly true that it is safe to use improper priors, specifically when posterior distributions are well-defined as limits of a sequence of approximating proper priors. However, when comparing models, improper parameter priors make Bayes factors indeterminate.

For this reason, we use the marginal posterior probabilities for k_1 and k_2 to do model inference. In fact, rather than selecting the “best” model, these probabilities allow the Bayesian to use model averaging strategies or more qualitative model comparisons.

4 Predictive densities

By using the BD-MCMC algorithm we can produce samples from the posterior distributions of the service and re-service times distribution. Thus, given the BD-MCMC output of size N after a burn-in period, for $\theta_{\sim 1}$ and $\theta_{\sim 2}$, and suitable regularity conditions [see, e.g., Tierney 1996, p. 65], these quantities of interest can be consistently estimated by the sample path averages. We first estimate the mixture size of service and re-service distribution, k_1 and k_2 . The estimates of the marginal posterior distributions of k_1 and k_2 are

$$\begin{aligned} \Pr(K_r = k | data) &= \lim_{N \rightarrow \infty} \frac{1}{N} \# \left\{ n : k_r^{(n)} = k \right\} \\ &\approx \frac{1}{N} \# \left\{ n : k_r^{(n)} = k \right\}, \quad (N \text{ large}), \quad r = 1, 2 \end{aligned} \quad (12)$$

This probability provides a tool for determining the number of phases of service and re-service distributions.

We can estimate the predictive density of the service and re-service time distributions using

$$\hat{B}_r \left(t \left| \begin{matrix} t \\ s_1 \end{matrix} \right. \right) \approx \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^{k_r^{(j)}} \pi_{ri}^{(j)} G \left(t \left| \alpha_{ri}^{(j)}, \beta_{ri}^{(j)} \right. \right), \quad r = 1, 2 \quad (13)$$

also, the predictive density of the service and re-service time distributions can be estimated by

$$\hat{B}_r \left(t \left| \begin{matrix} t \\ \theta_{\sim r} \end{matrix} \right. \right) = \sum_{i=1}^{\hat{k}_r} \hat{\pi}_{ri} G \left(t \left| \hat{\alpha}_{ri}, \hat{\beta}_{ri} \right. \right), \quad r = 1, 2$$

where \hat{k}_r has a maximum posterior probability in (12). Note that in the case where the posterior distribution of $k_r, \alpha_{\sim r}$ and $\beta_{\sim r}$ is fairly spread out or even multimodal, these plug-in estimates would give a poor approximation of the predictive densities.

5 Estimation of some performance measures via the BD-MCMC output

Given a sample realization of the MCMC output and a sample from $f \left(\lambda \left| \begin{matrix} t \\ t \end{matrix} \right. \right)$ and $f \left(p \left| \begin{matrix} x \\ x \end{matrix} \right. \right)$ of equal size, we can estimate performance measures. For example, given sample data, we would like to assess whether or not the system is stable. The system is stable if and only if the traffic intensity, ρ , is less than one. Thus, the estimation of the probability of having a stable system is

$$\Pr \left(\rho < 1 \left| \begin{matrix} t, x, s_1, s_2 \\ \sim \sim \sim \sim \end{matrix} \right. \right) \approx \frac{1}{N} \# \left\{ \rho^{(n)} < 1 \right\} \quad (14)$$

where according to (1) we have

$$\rho^{(n)} = \lambda^{(n)} \left(\sum_{i=1}^{k_1^{(n)}} \pi_{1i}^{(n)} \frac{\alpha_{1i}^{(n)}}{\beta_{1i}^{(n)}} + p^{(n)} \sum_{i=1}^{k_2^{(n)}} \pi_{2i}^{(n)} \frac{\alpha_{2i}^{(n)}}{\beta_{2i}^{(n)}} \right)$$

in which $\left\{ \left(k_1^{(i)}, \pi_{\sim 1}^{(i)}, \alpha_{\sim 1}^{(i)}, \beta_{\sim 1}^{(i)} \right) \right\}_{i=1}^N$ and $\left\{ \left(k_2^{(i)}, \pi_{\sim 2}^{(i)}, \alpha_{\sim 2}^{(i)}, \beta_{\sim 2}^{(i)} \right) \right\}_{i=1}^N$ are the samples of size N obtained from the BD-MCMC algorithm, also $\{\lambda^{(i)}\}_{i=1}^N$ and $\{p^{(i)}\}_{i=1}^N$ are the samples of size N generated from the posterior distributions of λ and p , respectively. A consistent estimator of the traffic intensity is

$$E \left(\rho \mid t, x, s_1, s_2 \right) \approx E \left(\lambda \mid t \right) \frac{1}{N} \sum_{n=1}^N \left(\sum_{i=1}^{k_1^{(n)}} \pi_{1i}^{(n)} \frac{\alpha_{1i}^{(n)}}{\beta_{1i}^{(n)}} + p^{(n)} \sum_{i=1}^{k_2^{(n)}} \pi_{2i}^{(n)} \frac{\alpha_{2i}^{(n)}}{\beta_{2i}^{(n)}} \right) \quad (15)$$

where $E \left(\lambda \mid t \right) = (\xi + n_t) / (\psi + \sum_{i=1}^{n_t} t_i)$.

Moreover, by using the MCMC estimations of the parameters system, i.e. $\left(\lambda, \theta_{\sim 1}, \theta_{\sim 2}, p \right)$, we can estimate the $\rho_1, \rho_2, \rho, \delta_1, \delta_2$, the mean system size (4) and (5), the mean busy period (2) and the probability of idle period of the system (3), as you see in the simulation study below.

6 Simulations

This section illustrates the accuracy of the Bayesian methodology in two simulation examples of the M/G/1 queuing system with optional second service. In the first example we assume both real density of service and re-service times are mixture of Gamma distributions, mixture of two Gamma distributions for service and mixture of three Gamma distributions for re-service times. In order to test how our methodology deals with *model misspecification*, the second example considers a more complicated model, in which the true density of service is a mixture of two truncated Normal distributions and the true density of re-service is Log-Normal distribution. The R codes are available at <http://www.math.rug.nl/stat> under the *research* link and will soon be available as R-package.

6.1 Simulation study: mixture of Gammas

Without loss of generality, we assume that the inter-arrival rate, λ , is known and equal to 0.26 and probability of re-service, p , is also known and equal to 0.3. We consider

samples of 1,000 service data from mixture of two Gamma distributions as below

$$B_1(s) = 0.6G(12, 1) + 0.4G(3, 2).$$

Also, for the re-service, 1,000 data simulated from a mixture of three Gamma distributions as below

$$B_2(\tilde{s}) = 0.6G(100, 100/3) + 0.3G(200, 50) + 0.1G(300, 60).$$

We assumed a Poisson prior distribution for k_1 with parameter $\gamma = 2$ which is truncated in point 100, and for remaining parameters we assume $\phi_1 = \dots = \phi_k = 1$, $\nu = \bar{s}_1^2/\sigma_{s_1}^2$, $\nu = 1/\sqrt{\nu}$, $\eta = (\bar{s}_1/\sigma_{s_1}^2)^{2/3}$ and $\tau = 1/\sqrt{\eta}$ in (7), (8) and (9), respectively. For the re-service data we take $\gamma = 3$ and the other parameters the same as their service equivalents.

For a service and re-service data set, we carried out the Bayesian approach described in Sect. 3.2. We ran 200,000 iterations of the BD-MCMC algorithm with 100,000 iterations as burn-in. From the diagnostics it is clear that these numbers exceed what is needed for reliable results. Methods for choosing a burn-in time and number of iterations to use after burn-in are discussed in Gilks et al. (1996).

Figure 1 provides the histograms of generated data set with the estimation of the predictive densities from formula (13) for service and re-service times. This figure shows that the productive densities for service and re-service times compare quite well with the true densities.

Figure 2 (left) illustrates the mixing properties of the algorithm in terms of the evolution of the mixture size k_1 and (right) is for k_2 . An essential element of the performance of our BD-MCMC algorithm is its ability to move between different values of k_1 and k_2 . The chains appear to be mixing quite well, visiting many states, for both service and re-service times.

Figure 3 on the left shows the estimation of posterior distribution of k_1 , which is obtained from formula (12), and on the right for k_2 .

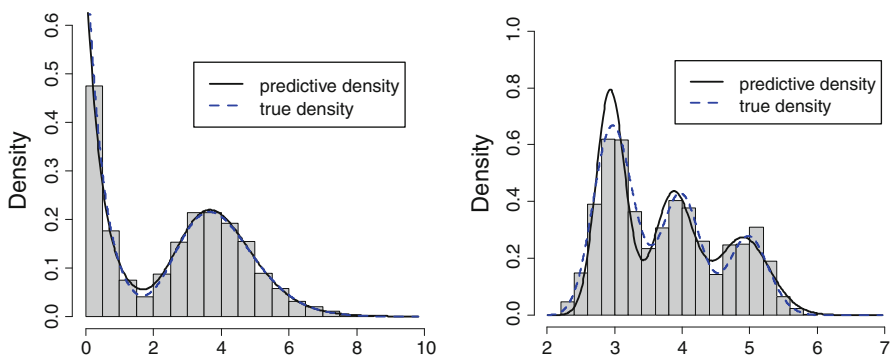


Fig. 1 Predictive densities (solid line) and the true densities (dotted line) for (left) service time data and (right) re-service time data

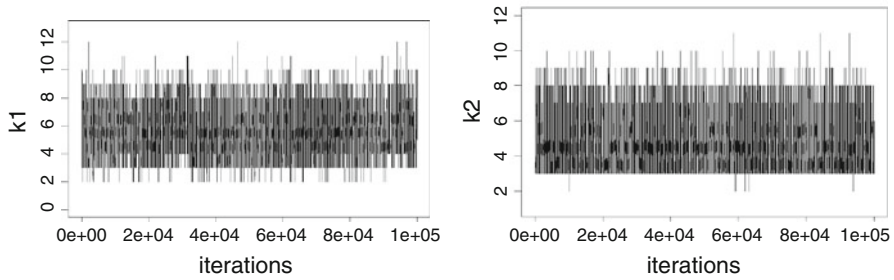


Fig. 2 (Left) A trace of k_1 for 100,000 iterations after 100,000 burn-in iterations and (right) for the k_2

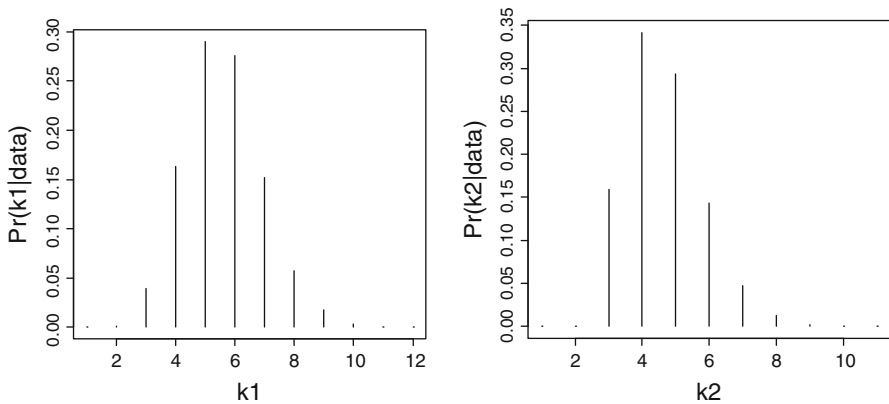


Fig. 3 (Left) The estimation of posterior distribution of k_1 , and (right) for k_2

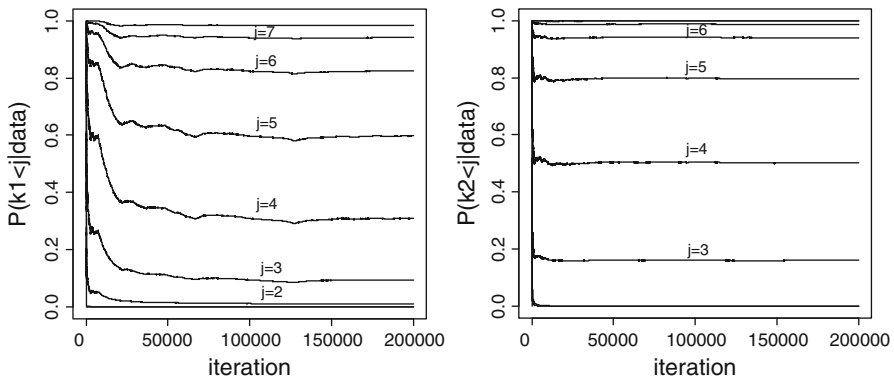


Fig. 4 The cumulative occupancy fractions for (left) service data, k_1 , (right) the re-service data, k_2 , for a complete run including burn-in

A useful check on the stationarity is given by the plot of the cumulative occupancy fractions for different values of k_1 and k_2 against the number of iterations. These are represented in Fig. 4 for the service and re-service data set, where it can be seen that the burn-in is more than adequate to achieve stability in the occupancy fractions.

Table 1 True values and estimations of mean busy period, probability of idle period, mean number of customers in the system, and probabilities that system is stable

Performance measures	E(busy period)	P(idleperiod)	$E(X_n)$	$E(Y_n)$	$P(\rho < 1 \text{data})$	$E(\rho \text{data})$
True value	9.0525	0.2982	1.4463	0.0029		0.9646
Estimates	9.036	0.2985	1.423	0.0028	0.9666	0.9666
SD	0.8207	0.017	0.1756	0.0004	0.0225	0.022

Third row is the standard deviation (SD) for these estimates

In the first and second row of Table 1, we respectively tabulate the true values and the estimated values of traffic intensity, probability of having a stable system, expectation of number of customers in MQ, expectation of number of customers in FQ, expectation of busy period and probability of idle period of the system. They have been obtained from formula (2), (3), (4), (5), (14) and (15), respectively. The third row of the table shows the standard deviation (SD) of these estimates.

Also, the real value of the traffic intensity, ρ , from formula (1) is equal with 0.952. A 95 % credible interval is roughly given as the estimate $\pm 1.96 \times \text{SD}$. Considering this criterion, it is concluded that all the true values lie inside their 95 % credible intervals.

6.2 Simulation study: mixture of truncated Normal and Log-Normal

In this section we would like to access the effect of model misspecification on our estimation procedure. Like in previous example, we assume that the inter-arrival rate and probability of re-service are known, $\lambda = 0.28$ and $p = 0.45$. We consider samples of 1,000 service data from mixture of two truncated Normal distributions on interval $(-\infty, 0)$ as below

$$B_1(s) = 0.4TN_{(0,\infty)}(1.4, 2.3) + 0.6TN_{(0,\infty)}(0.2, 0.3).$$

Also, for the re-service, 1,000 data simulated from a single Log-Normal distribution as below

$$B_2(\tilde{s}) = LN(1, 0.5).$$

With the same assumptions in previous example, we ran 200,000 iterations of the BD-MCMC algorithm with 100,000 iterations as burn-in.

Figure 5 shows the histograms of generated data set with real densities for service and re-service times and the predictive densities from formula (13). This figure shows the predictive densities for service and re-service times in comparison with the true densities. Despite the fact that the true model is not part of our model class, it is sufficiently rich to approximate quite general densities.

Figure 6 (left) illustrates the mixing properties of the algorithm in terms of the evolution of the mixture size k_1 and (right) is for k_2 . An essential element of the performance of our BD-MCMC algorithm is its ability to move between different values of k_1 and k_2 . The chains appear to be mixing quite well, visiting many states, for both service and re-service times.

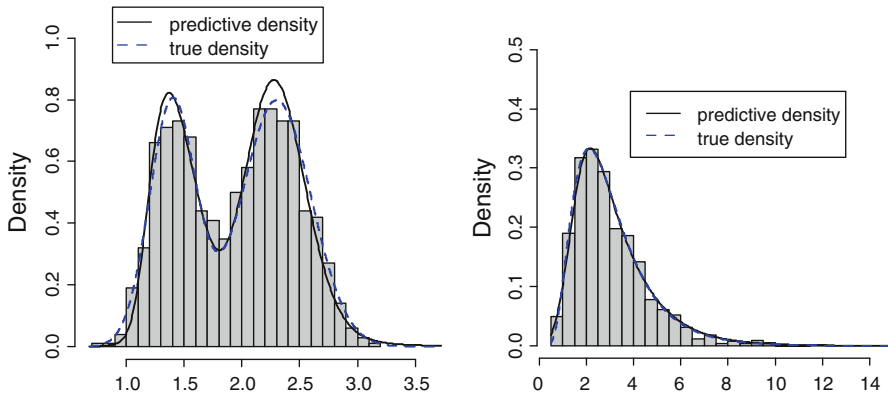


Fig. 5 Predictive densities (solid line) and the true densities (dotted line) for (left) service time data and (right) re-service time data

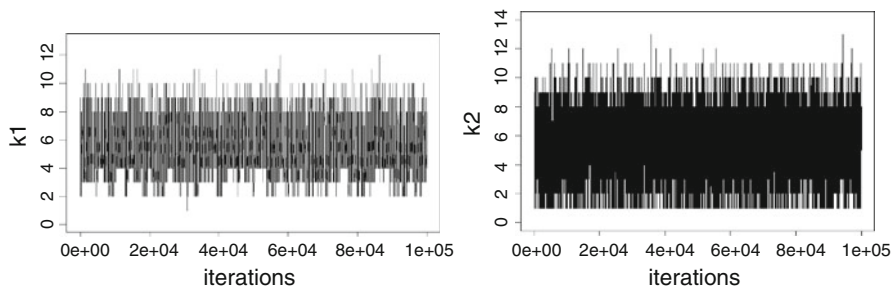


Fig. 6 (Left) A trace of k_1 for 100,000 iterations after 100,000 burn-in iterations and (right) for the k_2

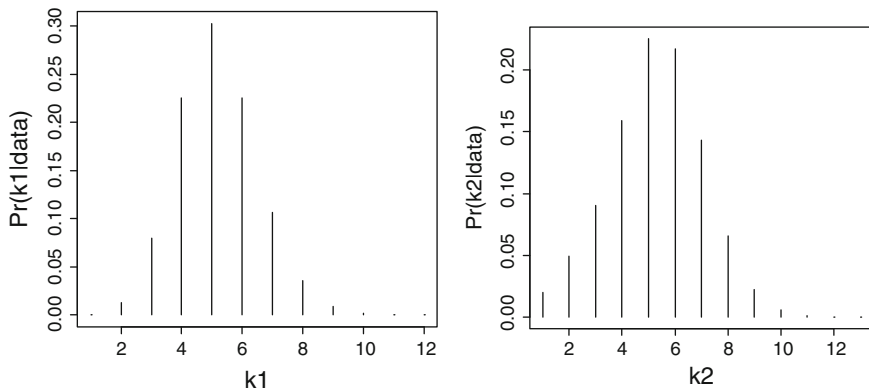


Fig. 7 (Left) The estimation of posterior distribution of k_1 , and (right) for k_2

Figure 7 in the left shows the estimation of posterior distribution of k_1 , which obtained from (12), and in the right is for k_2 .

For checking the stationary, Fig. 8 shows the cumulative occupancy fractions for different values of k_1 and k_2 against the number of iterations. These are represented

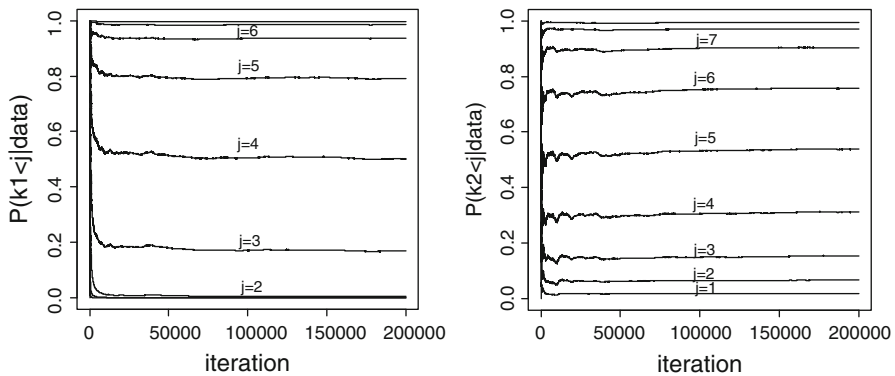


Fig. 8 The cumulative occupancy fractions for (*left*) service data, k_1 , (*right*) the re-service data, k_2 , for a complete run including burn-in

Table 2 True values and estimations of mean busy period, probability of idle period, mean number of customers in the system, and probabilities that system is stable

Performance measures	$E(\text{busy period})$	$P(\text{idle period})$	$E(X_n)$	$E(Y_n)$	$P(\rho < 1 \text{data})$	$E(\rho \text{data})$
True value	5.7974	0.381	1.661	0.227		0.983
Estimates	5.8445	0.3793	1.5863	0.2297	0.973	0.972
SD	0.0793	0.0032	0.0597	0.0024	0.0075	0.0075

Third row is the standard deviation (SD) for these estimations

for the service and re-service data set, where it can be seen that the burn-in is more than adequate to achieve stability in the occupancy fractions.

Table 2 respectively shows the true values and the estimated values of traffic intensity, probability of having a stable system, expectation of number of customers in MQ, expectation of number of customers in FQ, expectation of busy period and probability of idle period of the system.

Also $\rho = 0.983$, the real value of the traffic intensity from formula (1). Considering the values of SD, all the true values lie inside their 95 % credible intervals.

7 Discussion and future directions

This paper has developed a Bayesian approach to make inference and prediction for an M/G/1 queueing system with optional second service. It has developed a density estimation method based on a mixture of Gamma distributions in order to approximate the general service and re-service time distributions. A BD-MCMC algorithm has been proposed to make inference on the service and re-service parameters. This algorithm is based on births and deaths of mixture components making use of the birth-death technique proposed by Stephens (2000a). Some important measures of our queueing system, such as the system size mean, the mean busy period and probability of idle period have been estimated. This methodology has been illustrated with simulation study.

For our Bayesian approach, we used the mixture of Gamma distributions, since all distributions on the positive real can be approximated by such mixture and features such as skewness can be well-modeled by asymmetric distributions. There is previous work in Gamma mixture and Erlang mixture distributions (Wiper et al. 2001; Ausin et al. 2004, 2007; Ausin and Wiper 2007) but our Bayesian framework is different and its computation is easier and the results are more accurate. For instance, in the BD-MCMC algorithm, step 4.2., the Metropolis-Hastings step, in this article, the probability of acceptance is around 20 % which is more reasonable in comparison with previous work (e.g. see Ausin et al. 2004; Wiper et al. 2001). Furthermore, other family of distributions was used by Mohammadi and Salehi-Rad (2012), however Normal distribution is symmetric. A more suitable model would be a mixture of skewed truncated Normal distributions, which would make inference more complicated.

Our proposal is not limited to this queuing application, and comparisons with standard mixtures based on truncated normal distributions and skewed truncated normal distributions, see Frühwirth-Schnatter and Pyne (2010), are part of future work. Also, density estimation problems seem suitable for our approach. Work is currently in progress on these models.

An alternative to the BD-MCMC methodology is the RJ-MCMC methodology. This type of algorithm had been used by Wiper et al. (2001) for mixture of Gamma distributions and Ausin et al. (2004) for the mixture of Erlang distributions. In practice, we have found that both schemes perform similarly. In the BD-MCMC algorithm, as we have indicated, larger values of the birth rate produce better mixing, but also increase the computational cost. We have experienced some problems of non-convergence of the algorithm if the birth rate is selected too high. Thus, it would be useful to explore methods for selection of this parameter in order to optimize the algorithm.

Acknowledgments The authors are grateful to the anonymous reviewers for their detailed and insightful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Armero C, Bayarri MJ (1994a) Bayesian prediction in M/M/1 queues. *Queueing Syst* 15:401–417
- Armero C, Bayarri MJ (1994b) Prior assessments for prediction in queues. *Statistician* 43(1):139–153
- Ausin MC, Wiper MP (2007) Bayesian control of the number of servers in a GI/M/c queueing system. *J Stat Plan Infer* 137:3043–3057
- Ausin MC, Wiper MP, Lillo RE (2004) Bayesian estimation for the M/G/1 queue using a phase type approximation. *J Stat Plan Infer* 118:83–101
- Ausin MC, Wiper MP, Lillo RE (2007) Bayesian prediction of the transient behaviour and busy period in short- and long-tailed GI/G/1 queueing systems. *Comput Stat Data Anal* 52:1615–1635
- Cappé O, Robert CP, Rydén T (2003) Reversible jump MCMC converging to birth-and-death MCMC and more general continuous time samplers. *J R Stat Soc Ser B* 65:679–700
- Celeux G, Forbes F, Robert CP, Titterton DM (2006) Titterton deviance information criteria for missing data models. *Bayesian Analysis* 1(4):651–674
- Claeskens G, Hjort NL (2008) Model selection and model averaging. Cambridge University Press, Cambridge

- Diebolt J., Robert CP (1994) Estimation of finite mixture distributions through Bayesian sampling. *J R Stat Soc B* 56:363–375
- Frühwirth-Schnatter S (2006) Finite mixture and markov switching models. Springer, New York
- Frühwirth-Schnatter S, Pyne S (2010) Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics* 11(2):317–336
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. Chapman & Hall, London
- Green P (1995) Reversible jump MCMC computation and Bayesian model determination. *Biometrika* 82:711–732
- Green P (2003) Trans-dimensional Markov chain Monte Carlo. OUP, Oxford, pp 179–198
- Hastie D, Green P (2012) Model choice using reversible jump Markov Chain Monte Carlo. *Statistica Neerlandica* (Special issue, All models are wrong...)
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Medhi J (1982) Stochastic processes. Wiley Eastern, New Delhi
- Mohammadi A, Salehi-Rad MR (2012) Bayesian inference and prediction in an M/G/1 with optional second service. *Commun Stat Simul Comput* 41(3):419–435
- Ramirez P, Lillo RE, Wiper MP (2008) Bayesian analysis of a queuing system with a long-tailed arrival process. *Commun Stat Simul Comput* 4:697–712
- Richardson S, Green P (1997) On Bayesian analysis of mixtures with an unknown number of components. *J R Stat Soc B* 59:731–792
- Rios D, Wiper MP, Ruggeri F (1998) Bayesian analysis of M/Er/1 and M/Hk/1 queues. *Queueing Syst* 30:289–308
- Robert CP (1996) Mixtures of distributions: inference and estimation. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman and Hall, London, pp 441–464
- Salehi-Rad MR, Mengersen K (2002) Resourcing some customers in M/G/1 queues, under two disciplines, Advances in Statistics, Combinatorics and Related Areas. World Scientific Publishing, New Jersey, pp 267–274
- Salehi-Rad MR, Mengersen K, Shahkar GH (2004) Resourcing some customers in M/G/1 queues, under three disciplines. *Int J Math Math Sci* 32:1715–1723
- Sperrin M, Jaki T, Wit E (2010) Probabilistic relabeling strategies for the label switching problem in Bayesian mixture models. *Stat Comput* 20:357–366
- Stephens M (2000a) Bayesian analysis of mixture models with an unknown number of components an alternative to reversible jump methods. *Ann Stat* 28:40–74
- Stephens M (2000b) Dealing with label switching in mixture models. *J R Stat Soc B* 62(4):795–809
- Tierney L (1996) Introduction to general state-space Markov chain theory. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman & Hall, London, pp 59–74
- Wiper MP, Rios D, Ruggeri F (2001) Mixtures of gamma distributions with applications. *J Comput Graph Stat* 10:440–454